

# Analysing lexical cohesion in translation research using corpus linguistic methods

Olívia Péch

Eötvös Loránd University, Faculty of Arts  
Doctoral School in Linguistics, Translation Studies Programme  
Múzeum krt. 4/F, 1088 Budapest, Hungary  
[pech\\_oli@t-online.hu](mailto:pech_oli@t-online.hu)

**Abstract:** Lexical cohesion analysis is a subfield of the lexical semantic approach to text analysis, that is, the study of the semantic cohesion network as well as the supra-sentential organisation of texts. It examines the textually relevant repetition of lexical units. As languages exhibit varying degrees of tolerance to the use of cohesive devices (Halliday & Hasan 1976), it is argued here that Catford's notion of *shift* applies to translation regarding cohesion as well. This paper aims to explore whether translation between certain language pairs causes regular changes in the use of cohesive devices or not. It reports on the results of a lexical cohesion analysis, which was performed utilizing the advantages offered by computer-assisted corpus analysis. The study compares the number and the distribution patterns of cohesive devices characteristic of (authentic) source texts with those found in their translations (the target texts). I hope that through determining the text-level differences between authentic texts and translations my research will contribute to a better understanding of 'translationese' and thereby to raising the quality of translations.

## 1 Introduction

This paper has four objectives. First of all, I review those fields of study which are equally relevant to cohesion and translation research. I will also introduce a new approach to the study of translation exemplified by my own research. Then I move on to outline the advantages of computer-assisted research (which provides the technical background to this novel approach), as well as the obvious and hidden opportunities offered by corpus linguistic analysis, which in turn opens up new perspectives for empirical text analysis. Finally, I will present my hypothesis and research questions to be followed by the description of the method applied for lexical cohesion analysis and the analysis itself. The analysis compares the use of lexical cohesion devices in original Hungarian texts and their German translations using corpus linguistic methods.

## **2 Lexical cohesion research and translation studies**

### **2.1 The analysis of lexical cohesion**

Lexical cohesion analysis is a dominant field of text analysis, and is primarily concerned with the descriptive and functional analysis of written texts. The text-oriented approach began to gain popularity in the late 60s, as sentence grammar could not provide explanation for all the questions concerning linguistic phenomena at the supra-sentential (discourse) level (Tolcsvai Nagy 1994). For this school of thought, which began to treat text as a unified whole and as a basic unit of language, the emphasis is no longer on the fact that texts are made up of sentences. Instead, texts are essentially seen as being realized by sentences (Halliday & Hasan 1976).

Text semantics – the study of the semantic cohesion network – examines the repetition of textually relevant lexical units. By definition, “cohesion is a semantic relation between an element in a text and some other element that is crucial to the interpretation of it” (Halliday & Hasan 1976: 8). Lexical cohesion analysis classifies lexical elements into two functional groups. It focuses on central (relevant) lexical units, which convey the main ideas in the text, while lexical items loosely related or completely unrelated to the main topic are disregarded (Beaugrande & Dressler 2000).

### **2.2 Lexical cohesion research and translation studies: areas of joint research and results**

Even though translation studies has not become a discipline in its own right until as late as the second half of the past century, it has opened up to other fields quite early on. The nature of translation, that is, the task of encoding supra-sentential units of text makes a text-level analysis (e.g., observations about the use of cohesive devices as a result of translation from source to target language) absolutely necessary. Text-analytical approaches have two main focuses, whereby they represent both the descriptive categories of translation studies and the prescriptive requirements of teaching translation.

#### **2.2.1 Descriptive translation studies**

Several text linguistic studies have attempted to set the criteria for textuality before. A critical evaluation and comparison of these, however, falls outside the scope of this paper. For the purposes of the present study only those works will be reviewed which deal with text production, with special emphasis on those which look at lexical cohesion in the light of the translation activity.

While in a series of studies entitled *A fordító és a nyelvi normák [The translator and the linguistic norms]* Heltai classifies “the use of cohesive devices” as “belonging to the norms of text production” (Heltai 2004: 417), he also claims that “the norms of text production cannot be stated as explicitly as those of morphology or syntax.” Nevertheless, Heltai believes that “apart from conforming to the norms of prescriptive usage, which is in fact natural” (Heltai 2005b: 169), translators have to ensure

“optimal text cohesion” (Heltai 2005a: 36) and “during the construction of the text” they have to pay attention “primarily to text-level norms, conventions and distributional patterns”, part of which is the proper use of cohesive devices at both sentence and text level (Heltai 2005b: 169).

The functional significance of appropriate cohesion in translated texts is supported by Klaudy’s notion of quasi-correctness, which means that quite often “cohesion in translated Hungarian texts is degraded. The reader identifies sentence structures belatedly. The relationship between prominent and non-prominent units is not clear.” (Klaudy 1987: 7) In the case of quasi-correct texts, due to the deviation from the norms of target text production “the reader has to make an extra effort to be able to make out the meaning of an unusual linguistic form” (Klaudy 1987: 7). Moreover, as Klaudy observes, readers “refuse the text as a whole, because it does not conform to their intuitive notions of an adequately constructed text” (Klaudy 2004: 389).

Examining Hungarian translations prepared for a bilingual target audience living in Slovakia, Szabó Mihály presses for “a set of objective and unified criteria” to evaluate the translations against. Apart from having to adhere to the original “rhetoric aim; genre; grammatical and lexical register - terminology, technical language; thematic order”, appropriate text-level cohesion should also be a basic requirement (Szabó Mihály 2003: 58).

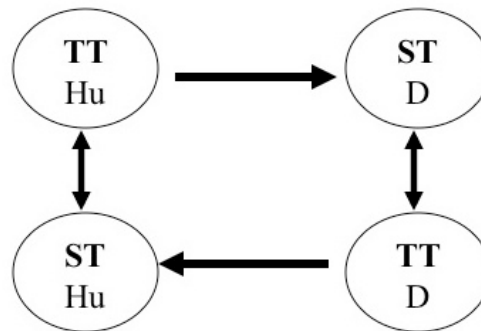
### **2.2.2 Teaching translation**

From the above it follows that the mastery of the appropriate use of cohesive devices in the target language is among the key objectives of teaching translation. The aim of teaching translation – as described in Dróth’s (2002) course description – “is to develop and extend skills needed to produce texts translated into one’s native language which do not sound strange or foreign at all and which are in accordance with the linguistic requirements and traditions of the target language” (95). Among the elements of translation competence to be gradually developed and refined during translator training, we can find “the ability to use cohesive devices properly” (Dróth 2002: 95). This competence area, to which considerable attention is paid during skill development, also features among the formative assessment criteria set in Dróth’s five-level error typology. Translated texts are classified and assessed based on the degree to which they come up to the norms and requirements of the target language, as “during the formative assessment of translations, the target text is not compared to the source text. Instead, the translation is analysed in isolation.” (Dróth 2002: 98)

## **2.3 A new approach put into practice**

Accepting the basic tenets of the literature outlined above, which hold that translation quality rests heavily on cohesion, the current research will focus on the analysis of changes in the use of cohesive devices as a result of translation. The essence of this new approach that emerged from the combination of translation studies and cohesion research is a comparative analysis. As the first step in the comparative analysis, I am going to examine the use of cohesive devices in original Hungarian texts and their German translations. At a later stage of this research project, I also plan to compare the use of cohesive devices in original German texts

and their Hungarian translations, and then, making use of all the previously analysed data, to perform a distribution analysis of lexical cohesive devices in original and translated texts in both Hungarian and German (Figure 1; abbreviations: ST: source text, TT: target text, Hu: Hungarian, D: German (from ‘Deutsch’))



**Figure 1.** Comparative analysis

### 3 Computer-assisted research

There are good reasons why comparative analyses of the sort envisaged above could not be performed earlier in the history of text analysis. Basically, comparative lexical cohesion analysis using traditional methods would be way too cumbersome and time-consuming. The traditional method is when the relations between a lexical unit and its various repetitions are indicated by lines connecting them (cf. Károly 2002: 80). It is easy to see that aligning sentences or complete texts this way would require superhuman efforts.

The sentence- and text-level comparison of a given source text and its target language equivalent was made available and conceivable thanks to the ever-more-popular corpus linguistic methods. Therefore, it is not an overstatement to say that it was the spread of new technology which made it possible to compare the lexical cohesion devices in translations with their source text realisations. The opportunities offered by corpus linguistic tools open up new perspectives for applied linguistic research both in terms of a smoother analytical process and additional query options.

#### 3.1 Structured encoding

The texts submitted to analysis were processed, that is, annotated with information on cohesive devices using the corpus linguistic software CLaRK. CLaRK, just as other software suited for similar linguistic tasks, allows for a structured encoding of texts to be machine-processed. This basically means that we can annotate the text with additional information (including the findings of linguistic analyses) while

preserving the integrity of the original text (Prószéky és Kis 1999). In the present study, I embedded the following types of meta information into the text: I marked sentence boundaries, selected and tagged recurring lexical units and different types of cohesive relations. For the texts opened in the CLaRK system to be visually clear and easy to follow, we can separate the original text from the annotated information by using colour-coding (Figure 2.).

```
<Text Analys="Sentence No Id">
  <Sentence No="NoId">
    <RepeatedLexUnit RLU="D.1.1">Die Personenschiffahrt</RepeatedLexUnit>

    <RepeatedLexUnit RLU="D.1.2">auf dem Balaton</RepeatedLexUnit>
begann im 19. Jahrhundert,
    <RepeatedLexUnit RLU="D.1.3">am 21 September 1846</RepeatedLexUnit>
  </Sentence>
```

**Figure 2.** Annotated excerpt

### 3.2 Query options

Machine-processing essentially means that we can run different queries on the same annotated source and detect different interrelationships every time. There are basically two query parameters: on the one hand, we have to define which (unit of) text we want to analyse and, on the other hand, we have to set the criteria for the analysis.

We can choose to study individual words, sentences, an entire source text and its translation or each and every analysed sentence in a corpus as well as their translations. It is exactly the free selection of any unit of the analysed text which makes the comparative analysis feasible in the first place. And the reason for this is that (word-, sentence- and text-level) alignment of the relevant units of the source and target texts is far less complicated and time-consuming for a machine than the same task carried out relying on human resources alone.

Considering the other query parameter: depending on the aims of our research, we can search the corpus against various criteria with the help of the meta-information supplied by the tags. We can query the exact number of tagged cohesive relations, their relative frequency, or even the context of the lexical cohesive devices. In other words, we can perform statistical and concordance analyses on the results (Balaskó 2005).

### 3.3 The advantages of machine-processing

As mentioned in the introduction, machine-processing was a key factor in making the comparative analysis happen. The advantages of applying corpus linguistic methods do not end with facilitating the comparative analysis; this framework is there to effectively help the linguist's analytical work in many other ways as well.

We can benefit from machine-processing during both the analysis phase and querying. Nevertheless, linguists have to make intellectual efforts to tag and analyse lexical cohesive devices as well as the relations between them – just like when these tasks are done the traditional way – as the annotation of a text with this type of information is not yet fully automated. It is of great help, though, that in the case of computer-assisted research, the structure of the text is adapted to the actual aims of the research, which may change or become broader in scope over time. Among the benefits of machine-processing, one can also mention the flexibility to integrate new research criteria either during or after the analysis. We can regroup the tagged data according to a new set of criteria at any stage. Moreover, the software also allows for the broadening or modification of the analytical categories. These custom settings can make our work even more efficient.

The wide range of available query options have already been introduced in the previous section. At this point, I only want to repeat that the software facilitates several different queries on the same annotated text. This, in turn, makes for an effective and seamless work-flow on one hand, and it facilitates the discovery of details and a refined analysis of causal relationships on the other. It is also important to note that the computer is a fast and reliable tool, which performs calculations and returns query results based on any criteria with a 100% accuracy. Additionally, the query results are readily accessible on computer. This is crucial for both storage and later use as one will not have to spend any more time on typing in the results obtained.

#### **4 Hypothesis**

The basic assumption underlying my research is that there is no one-to-one correspondence between the use of cohesive devices in different languages. If languages do exhibit varying degrees of tolerance to the repetition of cohesive devices, then Catford's (1965) notion of *shift* applies to translation in this respect as well. The results of my long-term research project will hopefully validate this hypothesis. The first step in attaining this goal, was to analyse the distribution of lexical cohesion devices in original Hungarian texts and their German translations.

At this point, two research questions will be explored: (1) To what extent is the use of cohesive devices preserved in a text translated from language A (source) to language B (target)? (2) Does translation between certain language pairs cause regular changes in the use of cohesive devices?

#### **5 Procedures of research**

##### **5.1 Choice of texts**

For my research I downloaded quasi-official, descriptive texts introducing different institutions, regions or organisations. The texts were chosen to meet three basic

criteria. (i) The style of the texts is similar due the fact that they are all descriptive, and this should effectively rule out the differences in the use of lexical cohesion devices found in texts of various styles due to pragmatic and stylistic reasons (Szikszainé Nagy 1999: 69). This is a crucial criterion, as “cohesion is a property which is only partly inherent to the text. The degree of cohesion varies from one type of text to another” (Vass 2002: 9). This factor has to be taken into account due to the frequency measurements and the comparability of the results thereof. (ii) The German translations were most probably done by professional translators, as these websites serve market interests. These high-quality texts guarantee that the dissimilarities in the use of cohesive devices in the analysed target texts are not due to the translator’s incompetence, but they are a result of the natural differences occurring in languages. (iii) Last but not least, downloading texts instead of typing them saved me a lot of time and effort.

## **5.2 Theoretical background**

The theoretical background of this research is provided by an analytical system worked out by Krisztina Károly in 2002, which looks into the text-organising function of lexical repetition. Károly’s taxonomy defines a lexical unit contributing to text cohesion as a unit whose meaning cannot be compositionally derived from the meaning of its constituent elements.

Károly distinguishes two basic types of cohesive relationships (in other words “ties”): lexical and text-based relations. With the first group divided into two subgroups, the following cases fall into distinct categories: (i) lexical units which recur in an unchanged form (repetition) and (ii) those instances in which it is only the information content of the lexical units which is repeated – either unmodified or modified, in the form of a semantically related lexical unit (synonymy, antonymy, hyponymy, meronymy). (iii) The text-bound relations form a category of their own (instantial relations). With respect to the type of repetition, synonymy or opposites, Károly makes a distinction between simple and derived forms. By derivation Károly refers to the cases when the lexical unit is repeated in a derived form (Károly 2002: 95-107).

Categories	Examples	
<b>LEXICAL RELATIONS</b>	<b>Simple</b>	<b>Derived</b>
<b>I. Same unit repetition</b>		
1. repetition	<i>Writers</i> do not hesitate. <i>Writers</i> communicate thoughts.	<i>Writers</i> do not hesitate. They <i>write</i> down their thoughts.
<b>II. Different unit repetition</b>		
2. synonymy	It's healthy <i>to exercise</i> every day. After <i>working out</i> at the fitness center I feel like a new man.	They've <i>built</i> several bridges this year. This <i>construction</i> company is efficient.
3. opposites	She <i>hates</i> him. Well, you can't <i>love</i> everyone.	What she now feels is <i>hatred</i> . But she used to <i>like</i> him.
4. hyponymy	Unfortunately, not all <i>birds</i> sing. But <i>cuckoos</i> do.	
5. meronymy	I can't use my <i>hands</i> . I've hurt my <i>fingers</i> .	
<b>TEXT-BOUND RELATIONS</b>		
6. instantial relations	The <i>teacher</i> laughed. I knew <i>Kitty</i> was a cheerful person. (equivalence) I tell you the story of my uncle's <i>cat</i> . His name was <i>Misty</i> . (naming)	

**Figure 3.** Károly's taxonomy of repetitions (Károly 2002:104)

Károly's (2002) taxonomy uses altogether nine categories of lexical cohesion to describe the syntactic-semantic network of the text-surface (Figure 3.). In my research, which compares Hungarian source texts and their German translations, I am going to apply these nine categories to explore the semantic relations crucial for text organisation.

### 5.3 Presentation of the analysis performed using corpus linguistic methods

As the first step of the analytical process, I marked the sentence boundaries and then numbered the sentences of the texts opened in the CLaRK system (Figure 4.). The added meta-information about sentence boundaries and the number of sentences can be interpreted by the software (our analytical tool) as well, which is essential for the sentence-level alignment of the original and translated documents. The letter 'd' together with a number (No1, No2) following it, stands for a particular German sentence, while the same number combined with letter 'h' stands for the corresponding Hungarian sentence.



```

-<Text Analys="Sentence No">
-<Sentence No="No1d">
  Die Personenschiffahrt auf dem Balaton begann im 19. Jahrhundert, am 21 September 1846.
</Sentence>
-<SentenceNo="No2d">
  An diesem Tage wurde das erste Dampfschiff "Kisfakudy" ins Wasser gelassen.
</Sentence>
-<SentenceNo="No3d">
  In den letzten 158 Jahren wurde unsere Firma zu einem bestimmenden Element des Tourismus am Balaton.
</Sentence>
-<SentenceNo="No4d">
  Unsere Gesellschaft ist die größte Schifffahrtsfirma auf dem See, ihre bekannteste Tätigkeit ist die Personenschiffahrt.
</Sentence>
-<SentenceNo="No5d">
  Der Großteil der Personenschiffahrt wird auf dem See von 26 Motorschiffen abgewickelt.
</Sentence>

```

**Figure 4.** Marking sentence boundaries and numbering the sentences

As the next step, I went on to analyse the first sentence for lexical cohesion, that is, to mark the lexical units relevant in terms of cohesive relations, most precisely, in terms of Károly's taxonomy of repetitions (Figure 2). These lexical units which will recur later on in the text were tagged with a (repeated lexical unit, RLU) reference number. The reference numbers D.1.1, D.1.2 and D.1.3 refer to the 1st, 2nd and 3rd lexical unit in the first sentence of the German text, respectively. The repeated lexical items in the rest of the sentences were tagged using the same logic. Using the reference numbers, I could connect the repetitions with the lexical units in the original sentences.

As the next step, I looked up all the repetitions of the lexical units selected from the first sentence. These were then analysed with respect to the type of cohesive relation and for each relation I assigned the relevant category of Károly's taxonomy (simple or derived repetition, simple or derived synonymy, simple or derived opposites) (Figure 5). The type of cohesive relation is followed by the reference number, which points to the first instance of the lexical unit (the repeated lexical unit).

The analysis of the rest of the sentences followed the same pattern: for every sentence I selected the lexical units and looked at their repetitions in the rest of the text. This way, I arrived at a complete analysis of the text, that is, each sentence was compared against all the other sentences in the rest of the text.

```

<Sentence No="No4d">Unsere Gesellschaft ist die größte
  <LexicalUnit CohesiveRelation="InstantialRelation" reference="D.1.1">Schifffahrtsfirma</LexicalUnit>
  <LexicalUnit CohesiveRelation="SimpleSynonymy" reference="D.1.2">auf dem See</LexicalUnit>
  , ihre bekannteste Tätigkeit ist
  <LexicalUnit CohesiveRelation="SimpleRepetition" reference="D.1.1">die Personenschiffahrt</LexicalUnit>
</Sentence>

```

**Figure 5.** Marking cohesion relationships

#### 5.4 The current state of the research

So far I have analysed three Hungarian texts and their German translations. The length of the texts varies between eleven to seventeen sentences, which also determines the number of analyses for each text. Since a single analysis involves the comparison of lexical units from at least two sentences, the complete analysis of a text usually consists of 10-16 analyses, which equals the number of sentences in the given text minus one. To date, I have performed 82 such analyses, which means that I have looked at the lexical cohesive units as well as the cohesive ties for altogether 82 sentences.

#### 5.5 Queries

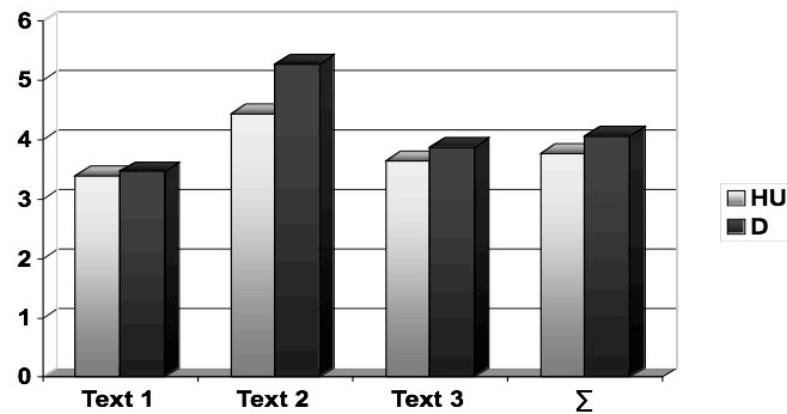
Considering that this research is still in the phase of corpus development and analysis, it would be too early to draw conclusions. The queries presented here are only meant to illustrate emerging tendencies.

I queried three complete texts, taking into account all the lexical units which are repeated – referred to – later on in the text (repeated lexical unit = RLU) as well as the repetitions themselves (lexical unit = LU). The results are included in the next table, broken down into texts and languages (Table 1.). Despite the almost equal number of sentences, I could detect marked differences between the texts.

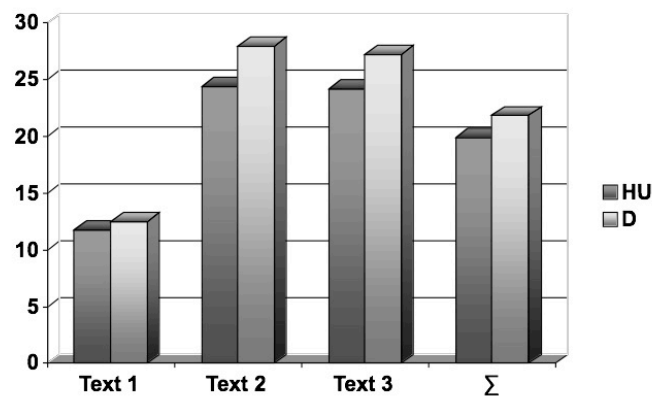
	Text 1	Text 2	Text 3
Hungarian			
Number of sentences	15	11	17
RLU	51	49	62
LU	177	268	412
German			
Number of sentences	17	11	17
RLU	59	58	66
LU	213	308	462

**Table 1.** Realisation of lexical repetition

Table 1 shows that based on the query on these three texts, the translated German texts have a higher frequency of lexical unit repetition than the Hungarian originals. This is visualised in the chart presented in Figures 6 and 7, which shows the repeated lexical units and the repetitions for each text and languages.

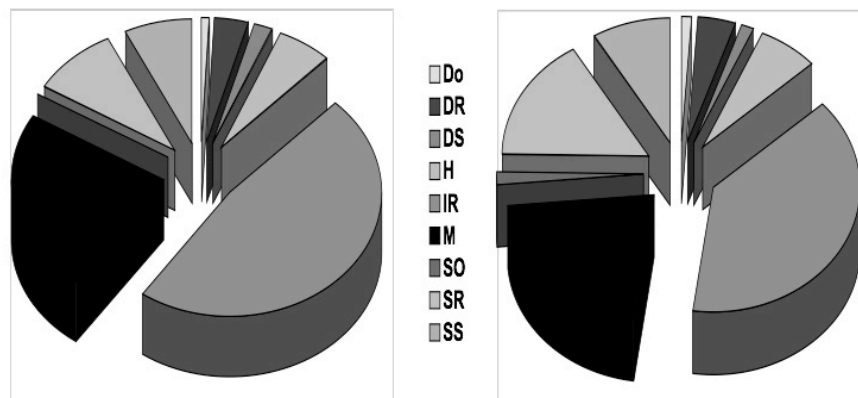


**Figure 6.** Frequency of repeated lexical units (RLU) by texts /total



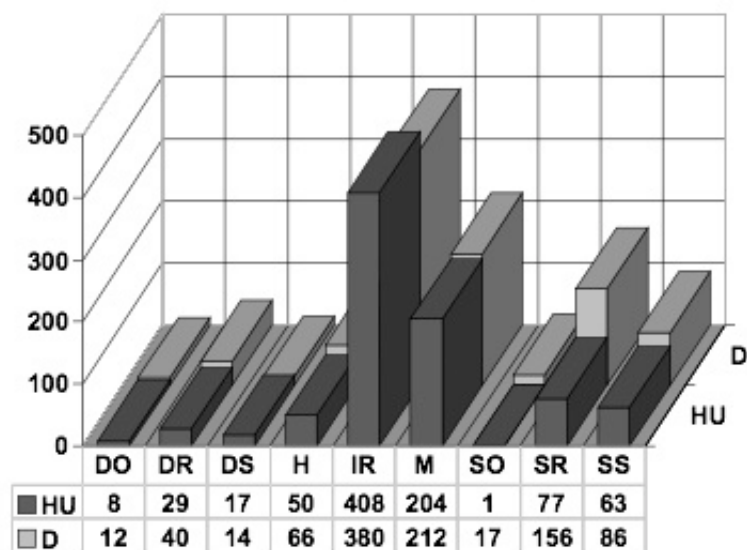
**Figure 7.** Frequency of repetitions (LU) by texts / total

One might infer from the dissimilarities presented above that the differences in frequency go hand in hand with those in the distribution of lexical cohesive devices in the two analysed languages. A visually-clear representation of the distribution is given in the pie chart below (Figure 8) (abbreviations: DO: derived opposites, DR: derived repetition, DS: derived synonymy, H: hyponymy, IR: instantial relations, M: meronymy, SO: simple opposites, SR: simple repetition, SS: simple synonymy).



**Figure 8.** The distribution of lexical units by languages (Hungarian-German)

For a more detailed picture, peruse the table and chart representing the distribution of lexical repetitions (Figure 9).



**Figure 9.** The distribution of lexical repetitions

While the query results should be handled with care due to the limited amount of data available, the tendencies beginning to show at the current stage of research hold the promise of confirming the hypothesis. Based on Table 1 – and at the current stage of data processing –, it seems that the translated German texts prefer the simple forms of repetition, synonymy and antonymy to the derived ones, which are more typical of the Hungarian source texts.

One could, of course, draw several other conclusions too from this set of data, but I would prefer to reflect on those after I have analysed a larger corpus of text. Moreover, space restrictions do not permit a presentation of the sentence- and word-level analysis of the query results. As outlined in the section on machine-processing, CLaRK has several other query options in store, which – in the later stages of research – will help unravel the exact causes behind certain dissimilarities.

## 6 Summary

The study presented above compares the lexical cohesion patterns in Hungarian source texts and their German translations. This comparative analysis is the fruit of a new interdisciplinary field at the intersection of text linguistics and translation studies. This novel approach – the comparative analysis of language pairs – was made feasible and conceivable by the ever-more-popular corpus linguistics methods, which facilitate the alignment of lexical units, sentences or excerpts from the source and target languages.

The tendencies emerging in the current stage of research suggest that the translated German texts show an increased frequency of cohesive devices and these texts prefer the simple forms of repetitions, synonymy and opposites as opposed to the Hungarian originals, which abound in derived forms.

Even though the current corpus size does not yet provide a full picture of dissimilarities between the cohesion patterns of Hungarian source texts and their German translations, the information extracted from the texts processed so far confirms the necessity of creating the two other sub-corpora mentioned in 2.3. And more importantly, the results confirm the necessity of those future analyses, which will focus on the comparison of original German source texts and their Hungarian translations on the one hand, and of original and translated texts in the same language on the other.

## Sources:

Művészetek Palotája. [Palace of Arts]

[http://www.mupa.hu/epulet\\_ahaz.jsp?language=hu&languageCode=0](http://www.mupa.hu/epulet_ahaz.jsp?language=hu&languageCode=0)

Palast der Künste. [http://www.mupa.hu/epulet\\_ahaz.jsp?language=ge&languageCode=2](http://www.mupa.hu/epulet_ahaz.jsp?language=ge&languageCode=2)

Magyar Fordítótárs. [Hungarian Translators House] <http://www.c3.hu/~bfordhaz/>

Das Ungarische Übersetzerhaus. <http://www.c3.hu/~bfordhaz/>

Balaton hajózás. [Balaton Shipping Co.] <http://www.balatonihajozas.hu/hu/rolunk.htm>

Balaton Schifffahrts AG. <http://www.balatonihajozas.hu/de/rolunk.htm>

## References

- Beaugrande, R. d., Dressler W. 2000. *Bevezetés a szövegnyelvészetbe. [Introduction to text linguistics]* Budapest: Corvina.
- Balaskó, M. 2005. *Korpusznyelvészeti vizsgálatok és fordításnyelvi minták (angol és magyar tudományos szövegek anyaga alapján). [Corpus linguistics analysis of samples from quasi-correct texts ('translationese') (based on technical texts in English and Hungarian)]* Pécsi Tudományegyetem: PhD dissertation
- Catford, J.C. 1965. *A Linguistic Theory of Translation*. Oxford: Oxford University Press.
- Dróth, J. 2002. Az értékelés mint rendszeres visszajelzés a fordítás oktatásában. [Evaluation as a regular feedback in teaching translation] *Porta Lingua 2002. Szaknyelvoktatásunk az EU kapujában. [Teaching Languages for Specific Purposes (LSP) in Hungary right before EU accession]* Debrecen: Debreceni Egyetem. 94-100.
- Halliday, M. A. K., Hasan R. 1976. *Cohesion in English*. Longman: London.
- Heltai, P. 2004. A fordító és a nyelvi normák I. [The translator and the linguistic norms I.] *Magyar Nyelvőr* Vol. 128. No. 4. 407-434.
- Heltai, P. 2005a. A fordító és a nyelvi normák II. [The translator and the linguistic norms II.] *Magyar Nyelvőr* Vol. 129. No. 1. 30-58.
- Heltai, P. 2005b. A fordító és a nyelvi normák III. [The translator and the linguistic norms III.] *Magyar Nyelvőr* Vol. 129. No. 2. 165-172.
- Károly, K. 2002. *Lexical repetition in text. A study of the text-organizing function of lexical repetition in foreign language argumentative discourse*. Frankfurt am Main: Peter Lang.
- Klaudy, K. 1987. Fordítás és aktuális tagolás. [Translation and information structure] *Nyelvtudományi értekezések* 123. Budapest: Akadémiai Kiadó.
- Klaudy, K. 2004. A kommunikatív szakaszhatárok eltűnése a magyarra fordított európai uniós szövegekben. [The loss of boundaries between communication units in EU documents translated into Hungarian] *Magyar Nyelvőr* Vol. 128. No. 4. 389-407.
- Nagy, Szikszainé I. 1999. *Leíró magyar szövegtan. [A descriptive grammar of Hungarian]* Budapest: Osiris.
- Prószéky, G., Kis B. 1999. *Számítógéppel emberi nyelven. Intelligens szövegkezelés számítógéppel. [Natural language and computers. Intelligent, computer-assisted text processing.]* Bicske: SZAK Kiadó.
- Szabó Mihály, G. 2003. A Szlovákiai magyar szakfordítások minőségének javításáról és az objektív fordításkritika megteremtésének feltételeiről. [On the quality improvement of Slovak-Hungarian technical translations and on the conditions of the creation of objective translation criticism] /English title by Szabó Mihály/ *Társadalomtudományi Szemle* Vol. V. No. 4. 55-68.
- Szirmai, M. 2005. *Bevezetés a korpusznyelvészetbe. A korpusznyelvészet alkalmazása az anyanyelv és az idegen nyelv tanulásában és tanításában. [Introduction to corpus linguistics. Application of corpus linguistic methods in teaching and learning first and foreign languages.]* Budapest: Tinta Könyvkiadó.
- Tolcsvai Nagy, G. 1994. *A szövegtan kialakulása. [The origins of text analysis]* Source: <http://www.jgytf.u-szeged.hu/~vass/szemml12.htm>
- Vass, L. 2002. *Előadásvázlatok. [Lecture notes.] Szövegtan. [Text analysis]* A szenzus. [Sensus - Sense] A rendszerszerű szenzus. [Sensus designatus] Diktum, apperceptum, evokátum. Kohézió. [Cohesion.] Kohézióhordozók. [Cohesive devices.] Source: <http://www.jgytf.u-szeged.hu/~vass/vsze111.htm>